Sequence, Evolution and Transcriptional Regulation of Avian-Mammalian P1 Type Protamines

Rafael OLIVA

Human Genome and Molecular Genetics Research Group,
Faculty of Medicine, University of Barcelona, Diagonal 643, 08028 Barcelona, Spain

ABSTRACT

Methodological approaches to protamine P1 sequence determination have evolved from the initial protein sequencing methods to the robust cloning and PCR-based techniques. Twenty-seven different mammalian-avian P1 type protamine genes and 32 different P1 amino-acid sequences are now available and allow detailed phylogenetic analysis and the study of transcriptional control mechanisms. All mammalian-avian P1 type protamines contain a well conserved N-terminus with the consensus "ARYR" followed by alternating "S/T-S-S" phosphorylatable residues. Eutherian mammalian P1s contain cysteine residues, whereas birds, prototherian and metatherian protamines lack cysteine. Thus cysteine appeared after the divergence of marsupials, monotremes and placental lineages. Overall detailed phylogenetic analysis of the gene sequences indicates that the evolution of P1 genes is in agreement with the expected species evolution supporting that these genes have evolved vertically.

RÉSUMÉ

Séquence, évolution et régulation transcriptionnelle des protamines de type P1 des Oiseaux et Mammifères

Les approches méthodologiques de détermination de séquence des protamines P1 ont évolué depuis les premières méthodes de séquençage de protéines jusqu'aux techniques fiables basées sur le clonage et la réaction d'amplification en chaîne. Vingt-sept gènes différents de protamines de type P1 des Oiseaux et Mammifères et trente-deux séquences différentes d'acides aminés sont maintenant disponibles et permettent une analyse phylogénique et une étude des mécanisme de contrôle de la transcription. Toutes les protamines de type P1 des Oiseaux et Mammifères contiennent une extrémité N-terminale bien conservée avec la séquence consensus "ARYR" suivie par la séquence alternée de résidus phosphorylables "S/T-S-S". Les protamines P1 des Mammifères Euthériens contiennent des résidus de cystéine, alors que les Oiseaux, les Prothériens et les Méthathériens n'en ont pas. La cystéine est donc apparue après la divergence des lignées des Marsupiaux, des Monotrèmes et des Placentaires. Une analyse phylogénique générale et détaillée des séquences de gènes indique que l'évolution des gènes des P1 est en accord avec l'évolution attendue des espèces, ce qui indique que ces gènes ont évolué verticalement.

Protamines are small (30-60 amino acids) and very positively charged proteins (40-70% arginine) which appear at the late stages of spermatogenesis in many but not all animal, and some plant, species [6, 11, 16, 21-23, 25, 43, 50, 52, 60, 69-73]. In those species in which they occur, such as in all mammals [6, 52], birds [13, 52, 53,], some teleost fish [11, 16], some reptiles [25, 70] and amphibians [25] they replace most of the histones during spermiogenesis and

OLIVA, R., 1995. — Sequence, evolution and transcriptional regulation of avian-mammalian P1 type protamines. *In*: JAMIESON, B. G. M., AUSIO, J., & JUSTINE, J.-L. (eds), Advances in Spermatozoal Phylogeny and Taxonomy. *Mém. Mus. natn. Hist. nat.*, **166** : 537-548. Paris ISBN : 2-85653-225-X.

become the major sperm nuclear protein [47, 48, 52, 54]. There are two basic groups of protamines in mammals; the P1 protamines which have been found in all mammalian species that have been analyzed and the P2 protamines which have been found in humans [5, 8, 42, 79, 80] and a limited number of other mammals such as mouse, guinea pig and stallion [10, 59, 66]. However, pro-P2 protamine genes have been sequenced from eight species of primates [66]. Both types of protamines contain cysteine which can form disulphide bonds and contribute to the stability of the condensed sperm nucleus. Bird protamines lack cysteine [13, 45, 49, 52] although they are clearly related to mammalian P1 protamines as several identical amino acid sequence motifs are present in both cases. Because of the high variability of protamines and protamine genes it is very difficult at present to explain the evolution of these proteins within an entire phylum. In many cases the limited number of sequences available precludes their connection into a coherent evolutive pathway. Thus the focus in this review has been placed in the avian mammalian-mammalian P1 type protamine for which a considerable amount of information is now available. Other papers and reviews cover other vertebrate or invertebrate groups ([4, 12, 13, 16, 25, 43, 49, 52, 60, 70, 71, 73], see also CHIVA, SAPERAS, CACERES & AUSIO, this volume, and PRATS & CORNUDELLA, this volume). Protamine genes are a clear example of highly tissue-specific genes. However the mechanisms that direct their specific expression in the testis are not fully understood [18, 22, 50, 52, 80]. Thus the last section of this review covers the progress made in the understanding of the transcriptional control of the P1 genes.

RESULTS AND DISCUSSION

Methodological approaches to protamine P1 sequence determination:

The methods initially available to sequence protamines were based on end group analysis, proteolytic digestion, isolation, sequencing, and overlapping of the protamine peptides. The presence of several arginine tracts in each protamine with very similar sequences made this approach technically difficult. The first reported avian-mammalian P1 type complete sequences using these methods corresponded to bull [15; Table 1], *Gallus domesticus* [45] and boar [76]. Some discrepancies in the initial reported sequences were found when the corresponding protamines were re-examined by automated micro-sequencing or by cloning of the protamine cDNAs and genes [37, 49, 40]. Subsequently the use of automated protein micro-sequencing led to the determination of the sequences of human P1 [41], stallion P1 [2, 9], ram [71] and rabbit, goat and rat [3]. Partial sequences corresponding to a few N-terminal residues have also been reported for many mammalian protamines [6] (Table 1).

Simultaneously to the onset of the use of automated micro-sequencing, the methods of the cDNA synthesis, clonning and sequencing were also developed and applied to protamine genes (Table 1). The first mammalian protamine cDNA sequence corresponded to mouse P1 [29]. Since no probes were initially available to screen the cDNA library, this initial sequence was obtained by characterization of selected clones preferentially expressed in spermatids [28]. Subsequently, the use of the mouse P1 cDNA clone as a probe led to the determination of the sequence of bovine protamine P1 cDNA [35]. Simultaneously, the bovine protamine P1 cDNA sequence was also independently obtained using oligonucleotides designed from the previously known amino acid sequence [31]. The mouse P1 cDNA also led to the isolation of the boar protamine 1 cDNA [37] and rat P1 cDNA [30]. The bovine probe led to the isolation and sequencing of the human protamine P1 cDNA [36]. However the mammalian probes would not recognize the avian protamine genes because of marked divergence in the nucleotide sequences between these species. Thus the cDNA sequence corresponding to rooster protamine was obtained by random sequencing of 210 clones from a rooster testis cDNA library until the sequence of one clone predicted an amino acid sequence similar to the previously reported at the protein level for galline [55]. The availability of a cDNA probe from galline led to the rapid isolation and sequencing of

TABLE 1. — Chronological list of reports on protamine P1 sequences with indication of the species and methods used for sequencing.

Year	Protamine sequence/s reported	Method	Reference
1972	Bull (*Bos taurus*)	Protein sequencing	[15]
1976	Galline (*Gallus domesticus*)	Protein sequencing	[45]
1976	Rat -partial-(*Rattus norvegicus*)	Protein sequencing	[27]
1983	Boar (*Sus scrofa*)	Protein sequencing	[76]
1984	Ram (*Ovis aries*)	Protein sequencing	[71]
1985	Human (*Homo sapiens*)	Protein sequencing	[41]
1985	Mouse (*Mus musculus*)	cDNA	[29]
1986	Bull (*Bos taurus*) (corrected)	Protein sequencing	[31]
1987	Bull (*Bos taurus*)	cDNA	[31]
1987	Bull (Bos *taurus*)	cDNA	[35]
1987	Stallion (*Equus caballus*)	Protein sequencing	[9]
1987	Stallion (*Equus caballus*)	Protein sequencing	[2]
1987	Human (*Homo sapiens*)	cDNA	[36]
1988	Galline (*Gallus domesticus*)	cDNA	[55]
1988	Bull (*Bos taurus*)	Genomic phage library	[32]
1988	Mouse (*Mus musculus*)	Genomic phage library	[24]
1988	Mouse (*Mus musculus*)	Protein sequencing	[10]
1988	Rabbit (*Oryctolagus cuniculus*)	Protein sequencing	[3]
1988	Goat (*Capra hircus*)	Protein sequencing	[3]
1988	Rat (*Rattus norvegicus*)	Protein sequencing	[3]
1988	Boar (*Sus scrofa*)	cDNA	[37]
1989	Chicken (*Gallus domesticus*)	Genomic cosmid library	[49]
1989	Quail (*Coturnix japonica*)	cDNA	[53]
1989	Rat (*Rattus norvegicus*)	cDNA	[30]
1989	Wallaby (*Cricetulus migratorius*)	Protein sequencing	[7]
1989	Dwarf Hamster (*Phodopus sungorus*)	Protein sequencing	[6]
1989	Rhesus monkey -partial-(*Macaca mulatta*)	Protein sequencing	[6]
1990	Human (*Homo sapiens*)	Genomic cosmid library	[18]
1991	Marmoset (*Saguinus imperator*)	PCR with genomic DNA	[63]
1992	Boar (*Sus scrofa*)	Genomic phage library	[26]
1993	Whale (*Orcinus orca*)	PCR	[1]
1993	Human (Mediterranean, Sudanese, Korean, American Indian)	PCR	[62]
1993	Rat -5´region-(*Rattus norvegicus*)	PCR	[64]
1993	Guinea pig -5´region-(*Cavia porcellus*)	PCR	[64]
1993	Gorilla -5´region-(*Gorilla gorilla*)	PCR	[64]
1993	Orangutan -5´region-(*Pongo pygmaeus*)	PCR	[64]
1993	Anubis baboon -5´region-(*Papio doguera*)	PCR	[64]
1993	Red monkey -5´region- (*Cercopithecus patas*)	PCR	[64]
1993	Opossum (*Didelphis marsupialis*)	PCR	[78]
1993	Common chimpanzee (*Pan troglodites*)	PCR	[68]
1993	Pygmy chimpanzee (*Pan paniscus*)	PCR	[68]
1993	Gorilla (*Gorilla gorilla*)	PCR	[68]
1993	Orangutan (*Pongo pygmaeus*)	PCR	[68]
1993	Gibbon (*Hylobates lar*)	PCR	[68]
1993	Red monkey (*Cercopithecus patas*)	PCR	[68]
1993	Marmoset (*Saguinus imperator*)	PCR	[68]
1993	Red howler (*Alouatta seniculus*)	PCR	[68]
1993	Platypus (*Ornithorincus anatinus*)	PCR	[67]
1993	Echidna (*Tachyglossus aculeatus*)	PCR	[67]
1994	Rat (*Rattus norvegicus*)	PCR	[61]
1994	Guinea pig (*Cavia porcellus*)	PCR	[61]
1994	Cat (*Felis catus*)	PCR	[61]
1994	Bear (*Ursus americanus*)	PCR	[61]
1994	Elephant (*Elephas*)	PCR	[61]
1994	Horse (*Equus caballus*)	PCR	[61]
1994	Camel (*Camelus dromedarius*)	PCR	[61]
1994	Deer (*Odocoideus virginianus*)	PCR	[61]
1994	Elk (*Cervus elaphus*)	PCR	[61]
1994	Moose (*Alces alces*)	PCR	[61]
1994	Gazelle (*Gazella dorca*)	PCR	[61]

another avian protamine, the quail (*Coturnix japonica*) [53], as well as to the isolation and sequencing of the chicken genomic clones [49]. The same approach using cDNA as probes to screen genomic libraries was also followed to obtain the genomic sequences corresponding to bull P1 [32], mouse P1 [24], human P1 [18] and boar P1 [26] (Table 1). The availability of probes for all these genes also led to the determination of their copy number which proved to be one copy per haploid genome for P1 protamines in mammals, two identical genes (coding region) per haploid genome in *Gallus domesticus*, and one copy of P2 per haploid genome in mammals. This indicated that the numerous basic protamine type molecules present in the sperm nucleus [P1, P2, P3, P4 and others] of mammals corresponded only to two types of protamines [P1 and P2]. The P1 genes are located adjacent to other spermatid-specific genes in the mammalian genome [19, 34, 46]. As some mammalian protamine genes were sequenced by independent laboratories, some minor discrepancies in the reported sequences also emerged, such as in the bull genes [31, 35] and between the human P1 sequence initially reported [18] with that subsequently redetermined in several independent individuals [62].

The availability of the P1 genomic sequences from human, bull, mouse and boar allowed their comparison in a search for conserved flanking sequences from which to design consensus oligonucleotides. This approach proved to be extremely efficient leading to the isolation and sequencing of the *Saguinus imperator* P1 protamine [63], *Orcinus orca* P1 [1], and subsequently several primates (common chimpanzee, pygmy chimpanzee, gorilla, orangutan, gibbon, *Cercopithecus patas*, *Alouatta seniculus*) [68], several human individuals (Mediterranean, Korean, Sudanese, American Indian) [62], additional eutherian mammals (rat, guinea pig, cat bear, elephant, horse, camel, deer, elk moose, gazelle) [61] and the monotremes, platypus and echidna [67]. The determination of a Wallaby partial amino acid sequence [7] by protein micro-sequencing led to the sequencing of the opossum protamine P1 [77, 78]. A similar PCR-based approach was followed to amplify and sequence the promoter region of the rat, guinea pig, gorilla, orangutan, anubis baboon and red monkey [64].

Although PCR from genomic DNA is a very valuable tool in the amplification and sequencing of new protamine genes it does not provide information on whether the sequenced genes are expressed or not (for instance, if a sequenced gene is a pseudogene). In the case of P1 genes, the fact that all mammalian species where the sperm nuclear protein content has been analyzed contain protamine P1 together with the single copy number of the P1 genes in mammals, suggests that most (if not all) of the mammalian species whose P1 sequence has been determined by PCR also express the sequenced gene. However this could be a limitation in the prediction of functional properties based on the derived amino acid sequence in the case of those proteins (such as P2 protamine) which are not ubiquitously expressed in mammals [66].

The availability, at present, of a large number of P1 sequences should allow design of new primers with which to amplify and sequence the P1 genes corresponding to species which have remained so far elusive. In the case of the P1 genes the PCR approach has been successful in the amplification of sequences corresponding to members of the class from which the oligos were predicted (e.g. mammals), but failed in the amplification of the P1 genes corresponding to other classes (e.g. reptiles or amphibians). Thus determination of the sequences of protamine genes in other vertebrate classes (or in other phyla) will probably require laborious groundwork based on either protein micro-sequencing (followed by oligonucleotide design and PCR) or cDNA based approaches [29, 55]. However once one or a few nucleotide sequences became available in the additional phyla and classes [4, see PRATS & CORNUDELLA, this volume], the same PCR-based approach successfully used in mammalian-avian P1s should also work in the determination of protamine sequences corresponding to most of the members of other classes.

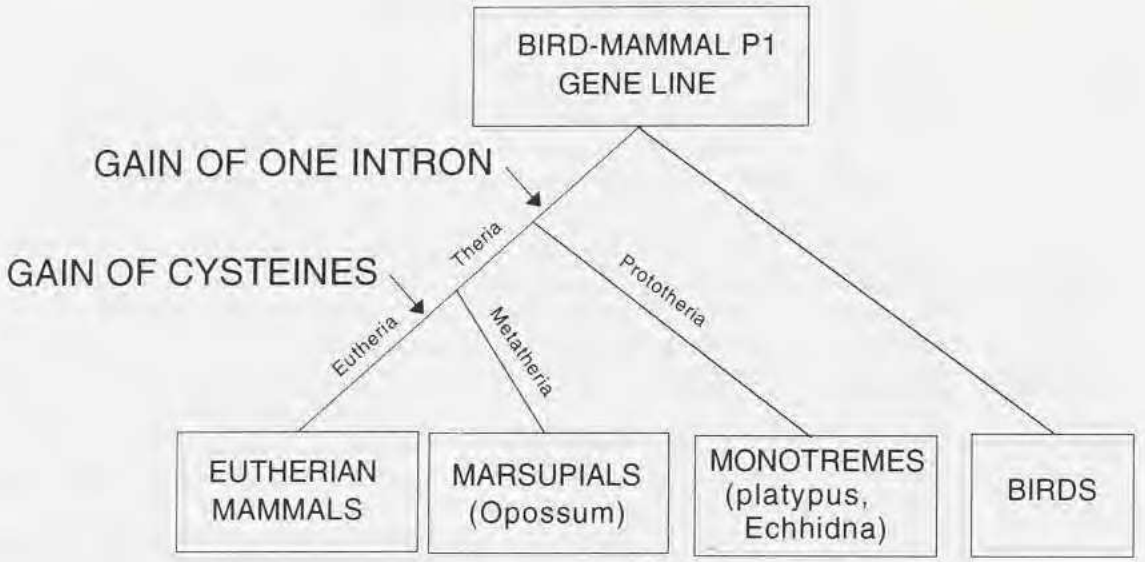
```
                10        20        40        40        50        60
Bull            ARYRCCLTH--SGSRCRRRRRRRCRRRRRR-FG------RRRRRR---VCCRR---YTVIRCTRQ
Goat            ARYRCCLTH--SRSRCRRRRRRRCRRRRRR-FG------RRRRRR---VCCRR---YTVVRCTRQ
Ram             ARYRCCLTH--SRSRCRRRRRRRCRRRRRR-FG------RRRRRR---VCCRR---YTVVRCTRQ
Orca            ARNRC-RSP--SQSRCRRPRRR-CRR-RIR-CC------RRQ-RR---VCCRR---YTTTRCARQ
Boar            ARYRCCRSH--SRSRCRPRRRR-CRRRRRR-CC------PRR-RRA--VCCRR---YTVIRCRRC
Horse           ARYRCCRSQ--SQSRCRRRRRRRCRRRRRR-SV------R-Q-RR---VSCRR---YTVLRCRRRR
Mouse           ARYRCCRSK--SRSRCRRRRR-RCRRRRRR-CC------R-RRRR---RCCRRRRSYT-IRCKKY
Rat             ARYRCCRSK--SRSRCRRRRR-RCRRRRRR-CC------R-RRRR---RCCRRRRSYT-FRCKRY
Rabbit          VRYRCCRSQ--SRSRCRRRRR-RCRRRRRR-CC------QRRRVR---KCCRR--TYT-LRCRRY
Pygmy chimp     ARYRCCRSQ--SRSRCYRQRR--SRRRKRQ-SC------QTQRRAM--RCCRRR--SR-LRRRRH
Common chimp    ARYRCCRSQ--SRSRCYRQRQ-RSRRRKRQ-SC------QTQRRAM--RCCRRR--SR-MRRRRH
Gorilla         ARYRCCRSQ--SRSRCYRQRQ-TSRRRRRR-SC------QTQRRAM--RCCRRR--NR-LRRRKH
Human           ARYRCCRSQ--SRSRYYRQRQ-RSRRRRRR-SC------QTRRRAM--RCCRPR--YR-PRCRRH
Orangutan       ARYRCCRSQ--SQSRCCRRRQ-RCHRRRRR-CC------QTRRRAM--RCCRRR--YR-LRCRRH
Red monkey      ARYRCCRSQ--SRSRCCRQRR-RCRRRRRR-RC------RARRRAM--KCCRRR--YR-LRCRRY
Gibbon          ARYRCCRSQ--SRSRCYRRGQ-RSRRRRRR-SC------QTRRRAM--RCCRPR--YR-LRRRRH
Red howler      ARYRCCRSRSLSRSRCYRQRP-RCRRRRRR-SC------RRP-RAS--RCCRRR--YR-LRRRRY
Marmoset        ARYRCCRSQ--SRSRCYRQRR-RGRRRRRR-TC------RRR-RAS--RCCRRR--YK-LTCRRY
Opossum         ARYRR-RSRSRSRSRYGRRRRR-SRSRRRR-SRRRRRRRG---RRG--RGYHRRSPHRRRRRRRRR
Echidna         ARFRPSRSR--SRSLYRRRRR--SRRQRSRRGGRQTGPRKITRRGRGRGKSRRRRGRRSMRSSRRRRRRRRN
Platypus        ARFRRSRSR--SRSLYRRRRR--SRR-----GGRQTRSRKLSR-SRRRGRSRRRKGRRSRRSSRRS--RRRN
Chicken         ARYRRSRTR--SRSPRSRRRRRRSGRRR---SPRRRRRYGSARRSRRSVGGRRRR-YGSRRRRRRRY
Quail           ARYRRTRTR--SRSR----RRRRSRRRR---SSRR-RRYGRSRRSYRSVGRRRRR-YGRRRRRRRRY
```

FIG. 1. — Alignment of the reported avian-mammalian P1 amino acid sequences (see Table 1).

Implications for protamine P1 gene evolution

The coincidence of the four C-terminal amino acids between mammalian P1s and galline (the protamine from *Gallus domesticus*) has been known for nearly two decades [45]. However, because of the lack of cysteine residues in bird protamines and the presence of these amino acids in mammalian protamines, both types of protamines had been classically classified as belonging to different types. According to BLOCH's classification, galline was type 1 (or a true protamine) whereas mammalian protamines were type 2 (stable or keratinous protamines). The similarity between mammalian and rooster protamine became stronger when the sequence predicted from the genome (and in accordance with the re-sequence obtained for the N-terminus of the protein) [49]; (Fig. 1) was used instead of the initial sequence reported [45]. For instance, the single threonine residue present in galline occupies exactly the same position as that present in ram and bull [52] (Fig. 1). The determination of the quail protamine sequence [53] also revealed the presence of the N-terminal ARYR motif and a size (56 residues) closer to mammalian P1s (50 residues) than those with galline (61 residues; Fig. 1). An alternating triple phosphorylatable site (T/S-S-S) is also found at positions 9-15 in all avian-mammalian P1 protamines (Figs 1, 2) [6, 20, 52, 57, 58]. The size among bird protamines appears consistently similar to that of mammals (Fig. 1) [13, 14]. Altogether the data strongly suggest the existence of an avian-mammalian protamine gene line during evolution (Fig. 2).

A new insight has come form the recent determination of the sequence corresponding to the prototherian (the monotremes platypus and echidna) [67] and the metatherian protamines (the marsupials) [7, 69, 77, 78]. All these sequences lack cysteine and the corresponding genes contain one intron. Thus these species are closer to birds according to their lack of cysteine but closer to eutherian mammals according to the presence of the single intron. Detailed phylogenetic analysis indicates that these sequences are half way between eutherian mammals and birds [7, 49, 67, 77, 78]. Based on the limited (but significant) similarity in the introns between prototherian



N-TERMINUS	ARYR	ARYR	ARFR	ARYR
T/S - S - S (9-15)	YES	YES	YES	YES
INTRON	YES	YES	YES	NO
CYSTEINES	YES	NO	NO	NO

FIG. 2. — Possible pathway of mammalian-avian protamine P1 gene evolution.

and eutherian mammals it was concluded that the introns in the protamine P1 genes of monotremes, marsupials and eutherian mammals were derived from a single intron that was probably inserted into the ancestral gene prior to the divergence of the theria and prototheria 150 to 170 million years ago [67] (Fig. 2).

The comparison of all protein and DNA sequences further strengthens the idea that protamines are amongst the most rapidly diverging proteins studied [68]. This variation may allow discrimination of closely related species or even individuals in some cases. For instance, a sequence polymorphism has been found in the human P1 gene [62]. Molecular analysis of the P1 genes from nine primates revealed that within primates the rate of evolutionary change is much higher than that within other mammalian orders [68]. Interestingly, the primate P1 data confirm that human-gorilla-chimpanzee P1 protamines are indeed very similar but, unlike the slightly closer association between chimpanzee and human derived from analysis of other genes [66], the human-gorilla relationship is slightly favoured in the case of the P1 genes [68].

Overall phylogenetic analysis of all P1 sequences (Table 1) indicates that the molecular evolution of P1 genes is in agreement with the expected species evolution supporting that these genes have evolved vertically [61, 83] (Fig. 2).

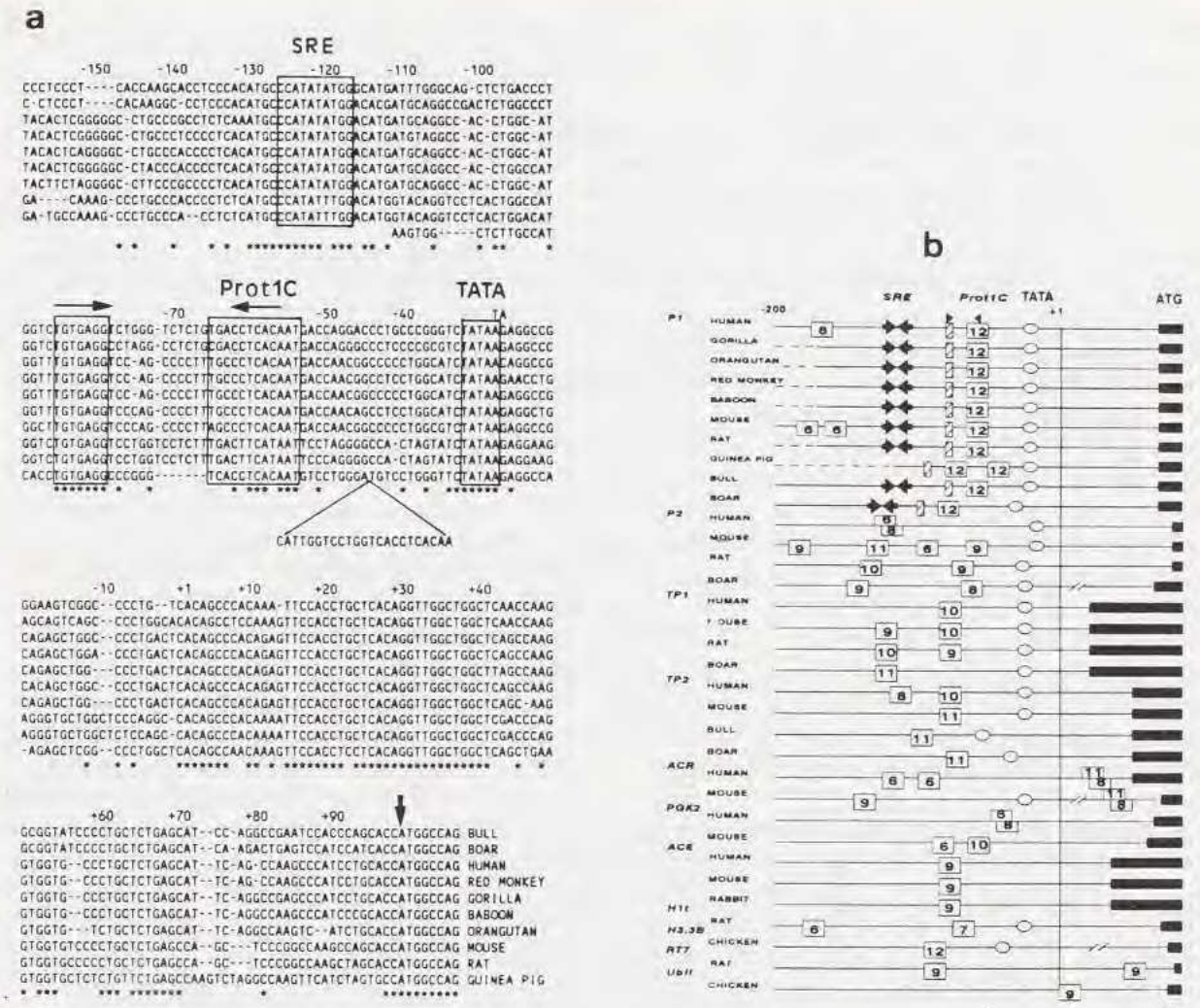


FIG. 3. — **a**: Alignments of all available mammalian *P1* gene promoter sequences. An asterisk indicates that a position is conserved in all species. The nt positions are referenced relative to the tsp [+1]. The conserved *SRE*, "TGTGAGG", *Prot1C* and the TATA box are boxed. The arrow over the *Prot1C* indicates the sequence which is palindromic with the "TGTGAGG" sequence (also arrowed) The start codon (ATG) is indicated by a downward arrow. After [64]. **b:** Position of the *Prot1C*, *SRE* and "TGTGAGG" sequences present in the *P1* genes and position of Prot1C-like sequences present in other testis-expressed genes. The putative *tsp* is indicated [+1]. The numbered open boxes indicate the position of sequences identical or similar to *Prot1C*. The number in the open box indicates the number of matches to the 12mer *Prot1C* [thus, 12 indicates a perfect match]. The beginning of the coding region is shown by the unnumbered solid boxes 3' to the tsp. A broken line indicates that the sequence is not available. The *SRE* is indicated by the two connected arrows facing each other. The "TGTGAGG" sequence is showed by a shaded box and the position and presence of a TATA box is indicated by the ellipse. After [64].

Transcriptional regulation of protamine P1 genes

Transcription of the avian-mammalian P1 type genes occurs in the post-meiotic, haploid stages of spermatogenesis as determined by Northern blot analysis of RNA from testis at different stages of development or from sorted cells [17, 35, 52, 55], and by *in situ* hybridization [35, 38, 44, 55]. Run-off assays on isolated mouse nuclei indicate that the mouse P1 gene is activated at the round spermatid stage [39]. What are the mechanisms leading to this specific activation? The availability of the nucleotide sequences from mouse, human and bull led to the prediction of some potential regulatory elements. For instance a TATA box is present in all of them [18, 24, 32], a

CRE-like element [50], CAT box, CG boxes and additional sequences [18, 24, 32, 50, 52]. A different approach in the prediction of potentially important sequences has been the comparison of homologous or heterologous protamine genes in the search for the conserved regions with the assumption that important regulatory sequences would have been conserved in evolution [50]. Thus the following comparisons were reported: mouse P1 and P2 genes [24]; human, bull and mouse P1 genes [33]; human P1 and P2 genes, mouse P1, bull P1 and human P1 [18]; porcine genes [26]; and chicken, bull P1, mouse P1 and trout protamine [50]. A common problem in all the studies comparing protamine gene P1 sequences was that the homology between the different P1 genes available for analysis was relatively high so that discrimination between conserved regulatory sites and sites conserved simply because of a close origin in evolution was not possible in many cases. This problem was solved when the promoter region of additional P1 genes was sequenced thus increasing the total number of sequences available for comparison [64] (Fig. 3).

Four highly conserved sites were detected in the 5´region (-160 to -1) of the protamine genes [64]. The first one (-29 to -35) corresponds to the already previously described TATA box, but with the novelty of being preceded invariably in all species by the di-nucleotide TC (Fig. 3). The second conserved region (-55 to -66) was named Protamine 1 Consensus (Prot1C) which is a CRE-like sequence (but always differently). The relevance of this conserved element in the expression of the P1 genes is strongly supported by the demonstration of a mouse testis trans-acting factor ([75], Tet-1) which binds and matches in the mouse the first 11 bp of the corresponding Prot1C sequence. Independent experiments using different oligonucleotides corresponding to the mouse P1 5´ region [82] showed that the region -35 to -70 led to the appearance of three different specific bands in gel retardation assays upon incubation with nuclear extracts from different tissues. Only one of those bands was testis-specific. Similar results have been obtained with the rat P1 sequence using rat nuclear extracts [65]. The third highly conserved region detected in the 5´of all P1 genes corresponds to the sequence TGTGAGG (-88 to -82). This sequence is a palindrome of the seven central nt of Prot1C (binding sequence for factor Tet1 in the mouse) and is exclusively present in this region in all P1 genes whose sequence is available suggesting an important function in the differential expression of the protamine P1 genes. This region corresponds to box E in the mouse promoter [24]. The fourth highly conserved region corresponds to MATGCCCATATWTGGRCAYG and has the typical structure of serum response elements (SRE) [64]. This region demonstrates specific interaction with a factor present in rat nuclear extracts from different tissues with a distinctive shifted band appearing in the testis extracts [65]. Also the equivalent region in the mouse P1 gene (described as Box O) demonstrates specific binding with a factor present in mouse nuclear extracts. Thus two highly conserved sequences, the Prot1C (also referred as to CRE-like in all P1 sequences or Tet1 binding site in the mouse), and the SRE appear to bind a testis specific factor or a testis-specific combination of factors [65]. Both of these sequences lie within the minimal region (from bp -150 to bp -37) described to direct spermatid-specific expression with a heterologous promoter from a human growth hormone reporter gene [56, 81, 82].

ACKNOWLEDGEMENTS:

This work was supported with grants from the "Fondo de Investigaciones Sanitarias" (FIS93/0670) from Spain, and from the Direccion General de Investigación Científica y Técnica (DGICYT PB92-0810).

REFERENCES

1. ADROER, R., QUERALT, R., BALLABRIGA, J. & OLIVA, R., 1992. — Nucleotide sequence of the protamine P1 gene from the whale *Orcinus orca* predicts a unique N-terminal amino-acid motif. *Nucleic Acids Research*, **20**: 609.

2. AMMER, H. & HENSCHEN, A., 1987. — The major protamine from stallion sperm. Isolation and amino acid sequence. *Biological Chemistry Hoppe Seyler*, **368**: 1619-1626.

3. AMMER, H. & HENSCHEN, A., 1988. — Primary structure of rabbit sperm protamine, the first protamine of its type with an aberrant N-terminal. *FEBS Letters*, **242**: 111-116

4. ARIYOSHI, N., HIYOSHI, H., KATAGIRI, C. H. & ABÉ, S.I., 1994. — cDNA cloning and expression of Xenopus sperm-specific basic nuclear proteins [SP5] gene. *Molecular Reproduction and Development*, **37**: 363-369.

5. ARKHIS, A., MARTINAGE, A., SAUTIÈRE, P. & CHEVAILLIER, P., 1991. — Molecular structure of human protamine P4 (HP4), a minor basic protein of human sperm nuclei. *European Journal of Biochemistry*, **200**: 387-392.

6. BALHORN, R., 1989. — Mammalian protamines: structure and molecular interactions. *In:* K. W. ADOLPH, *Molecular Biology of Chromosome Function*. New York, Springer Verlag: 366-395.

7. BALHORN, R., CORZETT, M., MAZRIMAS, J. A., CUMMINS, J. & FADEM, B., 1989. — Analysis of protamines isolated from two marsupials, the ring-tailed wallaby and gray short-tailed opossum. *Journal of Cell Biology*, **107**: 167a.

8. BALHORN, R., REED, S. & TANPHAICHITR, 1988. — Aberrant protamine 1 / protamine 2 ratios in sperm of infertile human males. *Experientia*, **44**: 52-55.

9. BÉLAICHE, D., LOIR, M., KRUGGLE, W. & SAUTIÈRE, P., 1987. — Isolation and characterization of two protamines St1 and St2 from stallion spermatozoa, and amino-acid sequence of the major protamine St1. *Biochimica et Biophysica Acta*, **913**: 145-149.

10. BELLVÉ, A. R., MCKAY, D. J., RENAUX, B. S. & DIXON, G. H., 1988. — Purification and characterization of mouse protamine P1 and P2. Amino acid sequence of P2. *Biochemistry*, **27**: 2890-2897.

11. BLOCH, D. P., 1969. — A catalog of sperm proteins. *Genetics Supplement*, **61**:93.

12. CÀCERES, C., RIBES, E., MULLER, S., CORNUDELLA, L. & CHIVA, M., 1994. — Characterization of chromatin-condensing proteins during spermiogenesis in a neogastropod mollusc [*Murex brandaris*]. *Molecular Reproduction and Development*, **38**: 440-452.

13. CHIVA, M., KASINSKY, H. E., MANN, M. & SUBIRANA, J. A., 1988. — On the diversity of sperm basic proteins in vertebrates: VI Cytochemical and biochemical analysis in birds. *Journal of Experimental Zoology* **354**: 404-317.

14. CHIVA, M., KASINSKY, H.E. & SUBIRANA, J. A., 1987. — Characterization of the protamines from four avian species. *FEBS Letters*, **215**; 327-240.

15. COELINGH, J. P., MONFOORT, C. H., ROZIJN, T. H., LEUVEN, J. A. G., SCHIPHOF, R., STEYEN-PARVE, E. P., BRAUNITZER, G., SCHRANK , B. & RUHFUS, A., 1972. — The complete amino acid sequence of the basic nuclear protein of bull spermatozoa *Biochimica et Biophysica Acta*, **285**: 1-14

16. DIXON, G. H., AIKEN, J. M., JANKOWSKY, J. M., MCKENZIE, D. I., MOIR, R. & STATES, J. C., 1985. — Organization and evolution of protamine genes of salmonid fishes. *In:* G. R. REEK, G. H. GOODWIN, & P. PUIGDOMENECH, *Chromosomal Proteins and Gene Expression*. New York, Plenum Press: 287-314.

17. DOMENJOUD, L., KREMLING, H., BURFEID, P., MAIER, W. M. & ENGEL, W. 1991. — On the expression of protamine genes in the testis of man and other mammals. *Andrologia*, **23**: 333-337.

18. DOMENJOUD, L., NUSSBAUM, G., ADHAM, I. M., GREESKE, G. & ENGEL, W., 1990. — Genomic sequences of Human protamines whose genes, PRM1 and PRM2, are clustered. *Genomics*, **8**: 127-133.

19. ENGEL, W., KEIME, S., KREMLING, H., HAMEISTER, H. & SCHLÜTER, G., 1992. — The genes for protamine 1 and 2 (PRM1 and PMR2) and transition protein 2 (TNP2) are closelly linked in the mammalian genome. *Cytogenetics and Cell Genetics*, **61**: 158-159.

20. GREEN, G. R., BALHORN, R., POCCIA, D. L. & HECHT, N. B., 1994. — Synthesis and processing of mammalian protamines and transition proteins. *Molecular Reproduction and Development*, **37**: 255-263.

21. HECHT, N. B., 1989. — Molecular biology of structural proteins of the mammalian testis. *In:* K.W. ADOLPH, *Molecular Biology of Chromosome Function*. New York, Springer Verlag: 396-420.

22. HECHT, N. B., 1990. — Regulation of haploid expressed genes in male germ cells. *Journal of Reproduction and Fertilization*, **88**: 679-693.

23. HECHT, N. B., 1993. — Gene expression during male germ cell development. *In*: C. DESJARDINS & L. L. EWING, *Cell and Molecular Biology of the Testis*. New York, Oxford University Press.

24. JOHNSON, P. A., PESCHON, J. J., YELICK, P. C., PALMITER, R. D. & HECHT, N. B., 1988. — Sequence homologies in the mouse protamine 1 and 2 genes. *Biochimica et Biophysica Acta*, **950**: 45-53.

25. KASINSKY, H. E., 1989. — Specificity and distribution of sperm basic proteins. *In*: L. S. HNILICA, G. S. STEIN & J. L. STEIN, *Histones and Other Basic Nuclear Proteins*. Boca Raton, Florida, CRC Press: 73-163.

26. KEIME, S., HEITLAND, K., KLUM, S., SCHLÖSSER, M., HROCH, N., HOLTZ, W. & ENGEL, W. 1992. — Characterization of four genes encoding basic proteins of the porcine spermatid nucleus and close linkage of three of them. *Biological Chemistry Hoppe-Seyler*, **373**: 261-270.

27. KISTLER, W. S., KEIM, P. S., & ANDERSON, R. L., 1976. — Partial structural analysis of the basic chromosomal protein of rat spermatozoa. *Biochimica et Biophysica Acta*, **427**: 752-757.

28. KLEENE, K. C., DISTEL, R. D. & HECHT, N. B., 1983. — cDNA clones encoding cytoplasmic poly(A)+ RNAs which first appear at detectable levels in haploid phases of spermatogenesis in the mouse. *Developmental Biology*, **98**: 455-464.

29. KLEENE, K. C., DISTEL, R. J. & HECHT, N. B., 1985. — Nucleotide sequences of a cDNA clone encoding mouse protamine 1. *Biochemistry*, **24**: 719-722.

30. KLEMM, U., LEE, C.H., BURFEIND, P., HAKE, S. & ENGEL, W., 1989. — Nucleotide sequence of a cDNA encoding rat protamine and the haploid expression of the gene during rat spermatogenesis. *Biological Chemistry Hoppe-Seyler*, **370**: 293-401

31. KRAWETZ, S. A., CONNOR, W. & DIXON, G. H., 1987. — Cloning of the bovine P1 protamine cDNA and the evolution of vertebrate protamines. *DNA*, **6**: 47-57.

32. KRAWETZ, S. A., CONNOR, W. & DIXON, G. H., 1988. — Bovine protamine genes contain a single intron. The structures of the two alleles. *Journal of Biological Chemistry*, **263**: 321-326.

33. KRAWETZ, S. A & DIXON, G. H., 1988. — Sequence similarities of the protamine genes: implications for regulation and evolution. *Journal of Molecular Evolution*, **27**:291-297.

34. KRAWETZ, S. A., HERFORT, M. H., HAMERTON, J. L., PON, R. T. & DIXON, G. H., 1989. — Chromosomal localization and structure of the human P1 protamine gene. *Genomics*, **5**: 639-645.

35. LEE, C. H., BARTELS, I. & ENGEL, W., 1987. — Haploid expression of a protamine gene during bovine spermatogenesis. *Biological Chemistry Hoppe-Seyler*, **368**: 807-811.

36. LEE, C. H., HOYER-FENDER, S. & ENGEL, W., 1987. — The nucleotide sequence of a human protamine 1 cDNA. *Nucleic Acids Research*, **15**: 7639.

37. MAIER, W. M., ADHAM, I., KLEMM, U. & ENGEL, W., 1988. — The nucleotide sequence of boar protamine 1 cDNA. *Nucleic Acids Research*, **16**: 11826.

38. MALI, P., SANDBERG, M., VUORIO, E., YELICK, P. C., HECHT, N. B. & PARVINEN, M., 1988. — Localization of protamine P1 mRNA in the different stages of the cycle of the rat seminiferous epithelium. *Journal of Cell Biology*, **107**: 407-412.

39. MATSUMOTO, M., KURATA, S., FUJIMOTO, H. & HOSHI, M., 1993. — Haploid specific activations of protamine 1 and hsc70t genes in mouse spermatogenesis. *Biochimica et Biophysica Acta*, **1174**: 274-278.

40. MAZRIMAS, J.A., CORZETT, M., CAMPOS, C., & BALHORN, R., 1986. — A corrected primary sequence for bull protamine. *Biochimica et Biophysica Acta*, **872**: 11-15.

41. MCKAY, D. J., RENAUX, B. S. & DIXON, G. H., 1985. — The amino acid sequence of human sperm protamine P1. *Bioscience Reports*, **5**: 383-391.

42. MCKAY, D. J., RENAUX, B. S. & DIXON, G. H., 1986. — Human sperm protamines. Amino acid sequences of two forms of protamine 2. *European Journal of Biochemistry*, **156**: 5-8.

43. MEZQUITA, C., 1985. — Chromatin proteins and chromatin structure in spermatogenesis. *In:* G. R. REEK, G. H. GOODWIN, & P. PUIGDOMENECH, *Chromosomal Proteins and Gene Expression*. New York, Plenum Press: 315-332.

44. MORALES, C. R., KWON, Y. K. & HECHT, N. B., 1991. — Cytoplasmic localization during storage and translation of the mRNAs of transition protein 1 and protamine 1, two translationally regulated transcripts of the mammalian testis. *Journal of Cell Science*, **100**: 119-131.

45. NAKANO, M., TOBITA, T., & ANDO, T., 1976. — Studies on a protamine (galline) from fowl sperm. 3. The total amino acid sequence of the intact galline molecules. *International Journal of Peptide and Protein Research*, **8**: 565-578.

46. NELSON, J. E. & KRAWETZ, S. A., 1993. — Linkage of human spermatid-specific basic nuclear protein genes. Definition and evolution of the P1-P2-TP2 locus. *Journal of Biological Chemistry*, **268**: 2932-2936.

47. OLIVA, R., BAZETT-JONES, D. P., LOCKLEAR, L. AND DIXON, G. H., 1990. — Histone hyperacetylation can induce unfolding of the nucleosome core particle. *Nucleic Acids Research*, **188**: 2739-2747.

48. OLIVA, R., BAZETT-JONES, D., MEZQUITA, C. & DIXON, G. H., 1987. — Factors affecting nucleosome disassembly by protamines in vitro. Histone hyperacetylation, time dependency and the size of the sperm nuclear proteins. *Journal of Biological Chemistry*, **262**: 17016-17035.

49. OLIVA, R. & DIXON, G. H., 1989. — Chicken protamine genes are intronless. The complete genomic sequence and organization of the two loci. *Journal of Biological Chemistry*, **264**: 12472-12481.

50. OLIVA, R. & DIXON, G. H., 1990. — Vertebrate protamine gene evolution I. Sequence alignments and gene structure. *Journal of Molecular Evolution*, **40**: 333-346.

51. OLIVA, R. & DIXON, G. H., 1991. — Expression and processing of the rooster protamine mRNA. *Annals of the New York Academy of Sciences*, **637**: 289-299.

52. OLIVA, R. & DIXON, G. H., 1991. — Vertebrate protamine genes and the histone-to-protamine replacement reaction. *Progress in Nucleic Acid Research and Molecular Biology*, **40**: 25-94.

53. OLIVA, R., GOREN, R. & DIXON, G. H., 1990. — Quail [*Coturnix japonica*], full length cDNA sequence, and the function and evolution of vertebrate protamines. *Journal of Biological Chemistry*, **264**: 17627-17640

54. OLIVA, R. & MEZQUITA C., 1986. — Marked differences in the ability of distinct protamines to disassemble nucleosomal core particles in vitro. *Biochemistry*, **35**: 6508-6511.

55. OLIVA, R., MEZQUITA, J., MEZQUITA, C. & DIXON, G. H., 1988. — Haploid expression of rooster protamine mRNA in the postmeiotic stages of spermatogenesis. *Developmental Biology*, **135**: 332-340.

56. PESCHON, J. J., BEHRINGER, R. R., BRINSTER, R. L. & PALMITER, R. D. 1987. — Spermatid-specific expression of protamine 1 in transgenic mice. *Proceedings of the Natural Academy of Sciences USA*, **84**: 5316-5319.

57. PIRHONEN, A., LINNALAKANKUNEN, A. & MAENPAA, P. H., 1994. — P2 protamines are phosphorylated in-vitro by protein dependent kinase-C, whereas P1 protamines prefer cAMP-dependent protein-kinase. A comparative study of 5 mammalian species. *European Journal of Biochemistry*, **223**: 165-169.

58. PIRHONEN, A., LINNALAKANKUNEN, A. & MAENPAA, P. H., 1994. — Identification of Phosphoseryl residues in protamines from mature mammalian spermetozoa. *Biology of Reproduction*, **50**: 981-986.

59. PIRHONEN, A., VALTONENE, P., LINNALA-KANKUNNEN, A., HEISKANEN, M. L. & MÄENPÄÄ, P. H., 1990. — Primary structures of two protamines 2 variants [St2a and St2b] from stallion spermatozoa. *Biochimica et Biophysica Acta*, **1039**: 177-180.

60. POCCIA, D., 1986. — Remodelling of nucleoproteins during gametogenesis, fertilization and early development. *International Review of Cytology*, **105**: 1-65

61. QUERALT, R., ADROER, R., OLIVA, R., RETIEF, J., WINKFEIN, R. J. & DIXON, G. H., 1995. — Evolution of Protamine 1 genes in mammals. *Journal of Molecular Evolution*, **40**: 601-607.

62. QUERALT, R., FÀBREGUES-BOIXAR, O., ADROER, R., GENÉ, M., GÓMEZ-CATALÁN, J., HUGUET, E. & OLIVA, R., 1993. — Direct sequencing of the human protamine P1 gene and application in forensic medicine. *Journal of Forensic Sciences*, **38**: 1491-1501.

63. QUERALT, R. & OLIVA, R., 1991. — Protamine 1 gene sequence from the primate *Saguinus imperator* isolated with PCR using consensus oligonucleotides. *Nucleic Acids Research*, **19**: 5786.

64. QUERALT, R. & OLIVA, R., 1993. — Identification of conserved potential regulatory sequences of the protamine-encoding P1 genes from ten different mammals. *Gene*, **133**: 197-204.

65. QUERALT, R. & OLIVA, R., 1995. — Demonstration of trans-acting factors binding to the promoter region of the testis-specific rat protamine P1 gene. *Biochemical and Biophysical Research Communications*, **208**: 802-812.

66. RETIEF, J. D. & DIXON, G. H., 1993. — Evolution of pro-protamine P2 genes in primates. *European Journal of Biochemistry*, **214**: 609-615.

67. RETIEF, J. D., WINKFEIN, R. J. & DIXON, G. H., 1993. — Evolution of the monotremes: the sequences of the Protamine P1 genes of Platypus and Echidna. *European Journal of Biochemistry*, **218**: 457-461.

68. RETIEF, J. D., WINKFEIN, R. J., DIXON, G. H., ADROER, R., QUERALT, R., BALLABRIGA, J. & OLIVA, R., 1993. — Evolution of protamine P1 genes in primates. *Journal of Molecular Evolution*, **37**: 426-434.

69. RETIEF, J. D., KRAJEWSKI, C., WESTERMAN, M., WINKFEIN, R. J. & DIXON, G. H., 1995. — Molecular phylogeny and the evolution of marsupial protamine P1 genes. *Proceedings of the Royal Society of London, B, Biological Series*, **259**: 7-14.

70. SAPERAS, N., AUSIO, J., LLORIS, D. AND CHIVA, M., 1994. — On the evolution of protamines in bony fish: alternatives to the "Retroviral horizontal transmission" hypothesis *Journal of Molecular Evolution*, **39**: 282-295.

71. SAPERAS, N., CHIVA, M. & AUSIÓ, J., 1992. — Purification and characterization of the protamines and related proteins from the sperm of a tunicate, *Styela plicata*. *Comparative Biochemistry and Physiology*, **103**: 969-974.

72. SAUTIÈRE, P., BÉLAÏCHE, D., MARTINAGE, A. & LOIR, M., 1984. — Primary structure of the ram [*Ovis aries*] protamine. *European Journal of Biochemistry*, **144**: 121-135.

73. SUBIRANA, J. A., 1982. — Nuclear proteins in spermatozoa and their interactions with DNA. *In*: J. ANDRÉ, *The Sperm Cell*. The Hague, Martinus Nijhoff: 197-213.

74. SUBIRANA, J. A., 1991. — Protein-DNA interaction in spermatozoa. *In*: B. BACETTI, *Comparative Spermatology 20 Years Later*. New York, Raven Press: 89-92.

75. TAMURA, T., MAKINO, Y. MIKOSHIBA, K. & MURAMATSU, M., 1992. — Demonstration of a testis-specific trans-acting factor Tet-1 in vitro that binds to the promoter of the mouse protamine 1 gene. *The Journal of Biological Chemistry*, **267**: 4327-4332.

76. TOBITA, T., TSUTSUMI, H., KATO, A., SUZUKI, H., NOMOTO, M., NAKANO, M. & ANDO, T., 1983. — Complete amino acid sequence of boar protamine. *Biochimica et Biophysica Acta* **744**: 141-146.

77. WINKFEIN, R. J., NISHIKAWA, S., CONNOR, W., OLIVA, R., & DIXON, G. H., 1991. — An opossum protamine gene. *Journal of Cell Biology*, **115**: 89a 518

78. WINKFEIN, R. J., NISHIKAWA, S., CONNOR, W. & DIXON, G. H., 1993. — Characterization of a marsupial sperm protamine gene and its transcripts from the North American opossum (*Didelphis marsupialis*). *European Journal of Biochemistry*, **215**: 63-72.

79. YEBRA, L. & OLIVA, R., 1993. — Rapid analysis of mammalian sperm nuclear proteins. *Analytical Biochemistry* **209**: 201-203.

80. YEBRA, L., BALLESCÀ, J. L., VANRELL, BASSAS, L. & OLIVA, R., 1993. — Complete selective absence of protamine P2 in Humans. *Journal of Biological Chemistry*, **268**: 10552-10557.

81. ZAMBROWICZ, B. P., HARENDZA, C. J., ZIMMERMANN, J. W., BRINSTER, R. L. & PALMITER, R. D., 1993. — Analysis of the mouse protamine 1 promoter in transgenic mice. *Proceedings of the Natural Academy of Sciences USA*, **90**: 5071-5075.

82. ZAMBROWICZ, B. P. & PALMITER, R. D., 1994. — Testis-specific and ubiquitous proteins bind to functionally important regions of the mouse protamine-1 promoter. *Biology of Reproduction*, **50**: 65-72.

83. ZOLZER, V. & VONHAGEN, H. O., 1995. — Amino-acid-sequences of P1 protamines and the phylogeny of eutherian mammals. A cladistic study. Comparative *Biochemistry and Physiology B - Biochemistry and Molecular Biology*, **110**: 805-815.